

System and method for data transfer between two or more connected software services

DESCRIPTION

CROSS-REFERENCES TO RELATED APPLICATIONS

[Para 1] This application claims the benefit of U.S. Provisional Application Serial No. 60/481,350, filed on September 10, 2003, entitled “Automatic, high-performance data transfer between two or more connected software services, such as but not limited to Web Services, with arbitrarily defined, and optionally complex input/output data structures”, which is incorporated herein by reference.

BACKGROUND OF THE INVENTION

[Para 2] 1. Field of the Invention

[Para 3] The present invention relates to the field of computer software and, in particular, to a system and method for automatic transfer of data and interface integrity enforcement between two or more linked software services, including but not limited to Web Services and semantic-based programming constructs, with arbitrarily defined and optionally complex input/output data structures.

[Para 4] 2. Description of the Related Art

[Para 5] In a service-oriented program, mapping and transferring data between the inputs and outputs of dependent software services, including but not limited to Web Services, running in a related sequence, is one of the most time-consuming software development tasks. In the current art, most of these mappings are manually coded in object-oriented programming languages such as C++, Java and C# and/or in syntax-based scripting languages such as XSLT. The manual coding of data transfer between related software services

results in hard-coded, and one-off solutions that are hard to create and time consuming to debug, and maintain. On the other hand, there are a few graphical mapping and automated data transfer tools used in the area of message-oriented application integration. These tools are best suited for mapping message documents with simple to average complexity of structure and limited nested plurality, and when applied to mapping optionally complex data structures resulting from input/output of software services, suffer from lack of many-to-many mapping functionality, performance and scalability. Furthermore, the current art is missing a high-performance mechanism for automatically enforcing data integrity for input and output data of a software service at runtime.

[Para 6] The current art of service-oriented programming and application development is lacking a specialized semantic-based means of mapping data from arbitrary data structures between the outputs and inputs of software services (as well as between semantic-based programming constructs) that has a scalable and high performance system for automatic transfer of data and integrity enforcement at runtime based on a semantic description. The absence of the said high performance, scalable mechanism that can apply to software services can seriously restrict the applications of service-oriented programming and architecture.

[Para 7] 3. General Background

[Para 8] A *software service*, or service for short, including but not limited to a Web service, is a discrete software task that has a well-defined interface and may be accessible over the local and/or public computer networks or maybe only available on a single machine. Web services can be published, discovered, described, and accessed using standard-based protocols such as UDDI, WSDL, and SOAP/HTTP.

[Para 9] A *software service interface*, in concept, represents the *inputs* and *outputs* of a black-boxed software service as well as the *properties* of that service, such as name and location. Take, for example, the interface of a simple software service named GetStockQuote, which retrieves simple stock quote information [FIGURE 1]. This service takes a ticker symbol input and

returns the last trade price amount as well as some additional stock quote details, such as the day high and day low. Note that in order to use, or *consume*, a service, only knowledge of its interface is required. This means that as long as the interface of a service remains the same, different implementations of the service can be swapped in and out without affecting its consumers. This, as well as the fact that a service is a language- and platform-neutral concept, is one of the keys to the flexibility of service-oriented architectures.

[Para 10] An *atomic service* is a software service that is implemented directly by a segment of software code. In the existing NextAxiom™ HyperService™ Platform, atomic Web services are dispatched via a library. A library is a light, language- and platform-neutral wrapper that is linked to one or more atomic Web service implementations. Atomic Web services are logically indivisible Web services that represent “raw materials” to the HyperService™ platform.

[Para 11] A *composite service* is a software service that consumes any number of other atomic or composite services. In the HyperService™ platform, a composite Web service is implemented with a metadata-driven model that is automatically interpreted by a high-performance run-time engine.

[Para 12] Visual metadata models, which represent composite software services implementations to the HyperService™ system, are created in a graphical, design-time environment and stored as XML models. This environment offers a new and powerful visual modeling paradigm that can be leveraged to enable the visual modeling of transactional behavior. This environment was specifically designed to enable collaborative, on-the-fly creation of software services by business process analysts or functional experts, who understand the business logic and application required to implement real-world business processes and applications, but have no knowledge of programming paradigms or Web service protocols. FIGURE 2 captures the implementation of a composite software service named “Expedite 3000 Series”. This service is used by a master planner to expedite 3000-series inventory items when they fall short on the shop floor. This service was

developed collaboratively and reuses services that were selectively exposed by the Inventory and Purchasing departments to the developers of this service.

[Para 13] Any software service that is consumed by a composite service model is said to be “nested” or “embedded” within that composite service. FIGURE 3 depicts a hypothetical composite service that resides in Chicago. This software service is composed of other composite services that are distributed across the country.

SUMMARY OF THE INVENTION

[Para 14] A principle object of the present invention is to provide a system and a method for defining and automating mapping of data between inputs and outputs of two or more connected software services, such as but not limited to “Web services”, with arbitrarily defined and optionally complex input/output data structures.

[Para 15] The present invention provides a means for storing data associated with inputs and outputs of a software service or to any complex data structure in memory. Furthermore, the present invention provides a set of methods for high-performance access to the stored data with optional plurality such that data can systematically be read and written at any depth of containment, regardless of its plurality. A mechanism is provided for enforcing the integrity of data while the data values are being set as well as after the last data values have been set. Furthermore, a semantic-based means, that does not require any coding or scripting, is provided through a mapping tool. The mapping tool is used to semantically define how data should be transferred between data elements. The mapping tool enforces a set of mapping rules, between inputs and outputs of the services, or more generally between two arbitrarily defined data structures. The said mapping definitions are expressed in metadata, using a markup language, for the use by the system that automates data transfer. The mapping definitions, together with the method used for storing and retrieving the data, allows the automating system to

systematically fetch data from one data structure and set data on another structure. An extensible framework for defining built-in transformation functions capable of transforming or operating on data, with well-defined input and output structures is provided. The said framework provides a method for transferring data from one data structure to the inputs of a built-in function and from the outputs of the built-in function to another data structure.

[Para 16] Other objects and advantages of this invention will be set in part in the description and in the drawings that follow and, in part, will be obvious from the description, or may be learned by practice of the invention. Accordingly, the drawings and description are to be regarded as illustrative in nature, and not as restrictive.

[Para 17] To achieve the forgoing objectives, and in accordance with the purpose of the invention as broadly described herein, the present invention provides methods, frameworks, and systems for defining and automating mapping of data between inputs and outputs of two or more connected software services or data structures, with arbitrary definition, and optionally nested, input/output data structures with optional plurality. In preferred embodiments, this technique comprises: using a mapping tool to define the relation and association between data elements of two or more arbitrarily defined data structures; enforcing a well-defined set of rules to restrict users as to what they can map; optionally, transferring data through built-in functions; storing the description of the said mapping using a machine readable format or markup language such as but not limited to XML; expressing the association and mapping of data elements in the said description using a unique path identifier with an absolute and relative path addressing scheme; using a key-based look-up molding technique for storing the actual data associated to connected data structures in memory; using a list, or equivalent structure, of lookup tables for storing the data and data type information where the said list corresponds to a Data Container with plurality; using the said lookup tables used to store the data associated with the data elements and data types; automating the transfer of the actual data based on

the said description of the mapping and the said method of in-memory storage of actual data and enforcing data integrity at runtime. Furthermore, techniques for storing and retrieving data associated to arbitrarily defined structures are disclosed that are also used by the automated mapping system of the present invention.

[Para 18] The techniques and methods of the present invention provide a high-performance and automated way of transferring data between any number of connected, arbitrary data structures and eliminates the time consuming and mundane task of manual coding or scripting of data transfer when creating software applications. In one embodiment of the present invention, semantic-based definition of data mapping and automated, high-performance and scalable data transfer from many-to-many software services, including but not limited to Web Services, enables rapid development of service-oriented applications. In general, service-oriented architecture and applications are known for their value in reducing the cost of integration and application development while extending the agility of an application. Without a high-performance, scalable and generic method for mapping, transferring and transforming data between the inputs and outputs of connected software services, service-oriented programming and architecture will become highly restricted in its applications and thus its benefits will become restricted to a subset of the problem domain that it could have otherwise been applied. The present invention addresses this shortcoming for the current art of service-oriented application development.

[Para 19] The present invention will now be described with reference to the following drawings, in which like reference numbers denote the same element throughout. It is intended that any other advantages and objects of the present invention that become apparent or obvious from the detailed description or illustrations contained herein are within the scope of the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[Para 20] FIGURE 1 graphically depicts the interface definition for a software service named 'GetStockQuote', including the definition of the data elements that comprise the service Inputs and Outputs.

[Para 21] FIGURE 2 shows the implementation of an example composite software service.

[Para 22] FIGURE 3 illustrates an exploded view of a composite software service that contains other software services, some of which are distributed in different locations.

[Para 23] FIGURE 4 shows a mapping tool that allows a user to map data elements between two connected software services, and thus define how data should be transferred from one service to the other during runtime by a system that automates the data transfer.

[Para 24] FIGURE 5 is an XML representation of the mapping generated by the mapping tool shown in FIGURE 4, where data elements are mapped between the outputs of a software service to the inputs of another.

[Para 25] FIGURE 6 shows a mapping between the two connected software services shown in FIGURE 4, but this time with a built-in function between the two software services. The mapping defines how to transfer data from the outputs of a software service to the inputs of a built-in function and from the outputs of the built-in function to the inputs of a software service.

[Para 26] FIGURE 7 is an XML representation of the mapping generated by the mapping tool shown in FIGURE 6, where data elements are mapped from the outputs of a service to the inputs of a built-in function and from the outputs of a built-in function to the inputs of a service.

[Para 27] FIGURE 8 shows the mapping of data elements from the outputs of a service to the inputs of a decision construct.

[Para 28] FIGURE 9 is an XML representation of the mapping generated by the mapping tool shown in FIGURE 7, where data elements are mapped from the outputs of a service to the inputs of a decision construct.

[Para 29] FIGURE 10 shows the mapping of hierarchical data elements with plurality that will result in the “flattening” of data as it is transferred from one service to another.

[Para 30] FIGURE 12 shows how data elements from one service can be mapped to many services as well as how data elements from many services can be mapped to one service.

[Para 31] FIGURE 13A shows data elements between two connected services can be mapped during design-time, with a mapping tool, where the mapping tool allows data elements of different types to be mapped when a logical conversion of data elements exist between the mapped types.

[Para 32] FIGURE 13B shows how data is automatically transferred from one service to another by a runtime system that automates data transfer and data conversion based on the mapping defined in FIGURE 13A.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

[Para 33] The performance shortcomings of the prior art are overcome and additional advantages in automating data mapping across arbitrarily defined data structures are provided through a method of storing, assigning and retrieving data elements that belong to arbitrarily defined data structures, given prior knowledge of the structure of the data. This method assumes that each data element is associated to a data structure and a description for each data structure is available. The description of a data structure must include the name, and optionally the type of each data element. The description must also accommodate a type of element, hereon referred to as a “Data Container”. A Data Container is a divisible data element, whose purpose is to contain other data elements. Furthermore, each type of element can have an associated Boolean attribute indicating the plurality of the expected data where one value for the attribute indicates that the data is singular and the other value indicates that the associated data may be plural. Additionally, a default value, and a Boolean attribute indicating whether the corresponding data value at

runtime is required or optional, may be associated to each data element where applicable. Other attributes associated with each data type may indicate data range, allowed data values, data format or other enforceable restrictions for each data type.

[Para 34] In one embodiment of the above method of storing and retrieving data, methods are provided to store and retrieve data associated to well-defined inputs and outputs of a software interface including, but not limited to, a Web service. Here, an in-memory representation of the description of the inputs and outputs of software service, described by WSDL (Web Service Description Language) in the case of Web services, is used to mold the signature of the data structures associated to the service inputs and outputs in a specially designed data component. The molded information includes, but is not limited to, the expected hierarchy and structure of the data, the name of the data elements, the type of the data elements, default values for the data elements and whether data values for the corresponding data types are required or optional at runtime.

[Para 35] The specially designed data Component is used to store data corresponding to an arbitrarily defined data structure that may be associated to the inputs or outputs of a software service. The data Component has a lookup table for storing the molded data type definition information, as well as the data defined by the data type definition.

[Para 36] The following unique technique is used for molding a data Component: A unique lookup table key is determined for each data element within a hierarchical structure by traversing the hierarchy of parent Data Containers: a) If all parent Data Containers are declared to be singular, the resulting key will contain a concatenation of the names of all the Data Containers, each name separated by any character not allowed as part of the name identifying the Data Container, concatenated with the name of the data element. b) Otherwise, each time a Data Container with plurality is crossed, a new object that holds a data structure containing a sequence of Components, hereon referred to as ComponentList, is instantiated and inserted in the lookup table of the last Component crossed with a key that is the result of

concatenation of the name of all the Data Containers so far, the variable representing the Data Containers so far is reset to the empty string, the last component crossed is set to the first component of the newly instantiated ComponentList, and the rest of the corresponding data structures is traversed applying the same logic until no more list Data Containers are crossed and the leaf data element is reached. Many variations of this step can be applied. For example, the molding of a path that crosses a ComponentList can be delayed at runtime until a corresponding Component needs to be inserted in the associated ComponentList. Note that a ComponentList corresponds to a Data Container with plurality according to the definition of the data structure.

[Para 37] Once a key corresponding to the path of a data element is molded in the lookup table of a Component, an appropriate marker object is inserted as the value of the key. For example, if the key corresponds to a traversal path of a plural Data Container that is required at runtime, the value of the key maybe a marker indicating: LIST_DATA_CONTAINER_REQUIRED. In an object-oriented language, the marker may be a static Object of a class defined for this purpose. In a procedural programming language, this marker may be identified by a unique integer value.

[Para 38] Based on the technique described above, all the leaf data elements (i.e. non-divisible data elements) in a hierarchical structure of data will be stored in the same Component (more precisely, the same lookup table of the Component) as long as all of the embedded data structures crossed in the path of accessing the data element are singular. A new Component (within a ComponentList) is only created when an embedded data structure is plural and the new Component corresponds to an entry of the plural data structure.

[Para 39] To reduce the system memory requirements and improve performance while accommodating multiple instances of the same molded data structure, the following technique can be used: a molded set of Components corresponding to a data structure is created and cached. Each time there is a request for molding another instance of that structure, a copy of the cached molded Component with all its contents is returned. In this way, all the keys of the molded Components can point to the same values and thus

save memory by the fact that separate storage is not used for the same key values.

[Para 40] To provide data access and modification methods, the molded data Component provides methods for setting and getting data values in correspondence to the traversal path according to the way the Component keys were populated. For example to set a singular data element, a `setAtomicData` method that takes a key and a value as arguments can be used. The user of the method needs to create a key corresponding to the relative path of singular Data Containers crossed for accessing the data element that is the concatenation of the name of all the elements crossed, separated with appropriate name separators. In order to cross plural Data Containers, the Component provides a method: `getComponentList` with an argument signifying the relative path for the Component List corresponding to the plural Data Container and according to the method of molding described. Once a handle to a `ComponentList` is obtained, the user can create new Components corresponding to the plural structure with a method `addComponent`, on the `ComponentList`, and then add data values to a specific component.

[Para 41] A concrete example is in order: consider a data structure defined as a singular Data Container named `Part`, and a data element under `Part`, named `Number`. To set the value '101' on the `Number` element, given a Component instance, `Comp`, that is molded based on the `Part` data structure, the user makes the following call: `Comp.setAtomicData("Part\name", "101")`. Now, consider a Data Container named `SalesOrder` with a plural Data Container structure under it named `Lines`; furthermore, consider the `Part` structure to be embedded under the `Lines` structure. Now, assume that a molded Component, `Comp`, is instantiated corresponding to the `SalesOrder` structure. In order to set the value '101' on the `Number` element, under the `Part` Data Container, first the `ComponentList` instance, say `CompList`, corresponding to `Lines` is retrieved by calling: `LinesCompList = Comp.getComponentList("SalesOrder.Lines");` then, a new Component corresponding to a single `SalesOrder` line is created, or the `n`th component by index would be retrieved. To add a new component, the following method can be called: `LineComp = LinesCompList.addComponent()`.

This method returns the new component and then the following method can be used to set the Number element on the Line Component:
`LineComp.setAtomicData("Part\Name", "101").`

[Para 42] For a person expert in the art, it is easy to add all methods required for setting and getting the value of plural data elements (such as `setAtomicListData`), or to set and get other specific data types. The important point here is the key-based retrieval and assignment method and the unique way of molding and traversing data components.

[Para 43] Through the use of markers, inserted at the time of molding the Components according to the definition of the underlying data structures, the method of the present invention can provide automatic enforcement of data integrity. For example, a `verifyAllRequiredDataAreSupplied` method can be provided on Component and ComponentList objects to verify that all required data has been supplied. If a required data element was never set, the presence of static markers signifying REQUIRED data, indicate that the mold marker was never overwritten and thus a particular data element with a known path was never set. By associating a type system to the in-memory representation of a data structure definition, additional information about the expected type of each data element can be molded within a Component. The type system can be used to enforce the integrity of data based on the declared types through the implementation of the methods provided for setting data on the Component object.

[Para 44] We now turn to FIGURES 4 through 9 to focus our attention on a method of defining data mappings between two arbitrarily defined structures and automating the actual transfer of data based on a description of the mapping. Shown in FIGURE 4, FIGURE 6 and FIGURE 8, a mapping tool enables the user to graphically map data elements from two data structures, associated with the inputs and outputs of two software services. The same mapping tool can be used to graphically map data elements from many-to-many software services, for example, FIGURE 12 illustrates a mapping between many services to one service and one service to many services. A unique description of the

mapping, supporting the automatic transfer of data using the method of the present invention is generated and stored for the runtime system.

[Para 45] FIGURE 5, FIGURE 7, and FIGURE 9, provide an XML representation of the mapping generated by the example shown in FIGURE 4, FIGURE 6, and FIGURE 8, respectively. In general, there are two sides to the mapping: one is the 'FROM' side, the source of data, and the other is to the 'TO' side, where the data is being transferred. The mappings are organized based on the TO side of the mapping in a hierarchical fashion. FIGURE 5, FIGURE 7, and FIGURE 9 show the ToListGroup XML tag that represents the beginning of the mapping definition. For each plural Data Container on the TO side, a ToListGroup is added to the containing ToListGroup that corresponds to the last plural Data Container crossed, or the root ToListGroup. Each ToListGroup element contains CPFromGroup tags that correspond to a relation between the root Data Container on the FROM side, or any plural Data Container on the FROM side. A CPFromGroup contains attributes containing the relative lookup key in the Component that contains the data on the FROM side corresponding to the structure of the data on the FROM side. Furthermore, the CPFromGroup element contains other information such as all the paths required for traversing Components corresponding to plural Data Containers that are embedded within other plural Data Containers as well as specific mapping within an element called a ConnectedPair. The ConnectedPair element contains the relative data path of the data element on the FROM side of the mapping and the relative data path of the data element on the TO side of the mapping as well as other information.

[Para 46] In the method of the present invention, the Components within a ComponentList on the TO side can only be driven by the number of Components within a single ComponentList corresponding to a CPFromGroup on the FROM side. One exception to this rule is displayed in FIGURE 10 and FIGURE 10, when more than one CPFromGroup drives the insertion of Components on the TO sides ComponentList where all the driving ComponentLists of the FROM side have a hierarchical relationship. In which case, the data from the FROM side is said to have been 'Flattened'. These rules

are enforced through the graphical mapping tool used to generate mapping metadata [FIGURE 10].

[Para 47] Using the method of storing and retrieving of the present invention coupled with the method of expressing the mapping metadata, the system of this invention can easily automate the transfer of data from many Components to many Components corresponding to arbitrarily defined data structures. In one embodiment of the present invention, the in-memory representation of the ToListGroup mapping structure can be recursively traversed while the root Component corresponding to the TO side is traversed, or created if it doesn't exist, parallel to the traversal of the corresponding ToListGroup. While visiting each ToListGroup, if a CPFromGroup exists, a ComponentList is added to the TO side, and for each Component from the ComponentList on the FROM side, a Component is added to the ComponentList of the TO side and the data corresponding to all the relative data paths are looked up from the FROM Component and transferred to the corresponding TO Component.

[Para 48] Based on the method of the present invention, inputs/outputs from one or more software services may be mapped to a data-driven programming construct such as a decision (or branching) construct. FIGURE 8 shows an example where the TO side of the mapping is a decision construct.

[Para 49] As illustrated in FIGURE 13A and FIGURE 13B, the method of the present invention provides for automatic data conversion during runtime when two data elements of different types are mapped to each other during design-time, and when there exists a logical conversion between the two types. For example, an Integer value of 0 can be converted to a Boolean value of 'false', whereas any non-zero Integer value can be converted to a Boolean value of 'true'. A mapping tool can prevent a user from mapping two data elements of different types when a logical conversion of the data is not possible upon transfer.